



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 148592

TO: Vanessa L Ford
Location: 3b25 / 3c18
Monday, March 28, 2005
Art Unit: 1645
Phone: 571-272-0857
Serial Number: 10 / 017168

From: Jan Delaval
Location: Biotech-Chem Library
Remsen 1a51
Phone: 571-272-22504

jan.delaval@uspto.gov

Search Notes

From: Chan, Christina
Sent: Wednesday, March 23, 2005 8:02 AM
To: Ford, Vanessa; STIC-Biotech/ChemLib
Subject: RE: In re: 10/017,168 Sequence search

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

RECEIVED
MAR 23 2005
STIC

-----Original Message-----

From: Ford, Vanessa
Sent: Tuesday, March 22, 2005 7:04 PM
To: Chan, Christina
Subject: In re: 10/017,168 Sequence search

Please search SEQ ID NOS: 1, 3,5,19, 21, 23 and 25. Please include interference searches. Please rush.
Vanessa L. Ford
Biotechnology Patent Examiner
Office: REM 3B25
Mailbox: REM 3C18
Phone: 571.272.0857
Art unit:1645

STAFF USE ONLY

Searcher: Chan
Searcher Phone: 2-2504
Date Searcher Picked up: 3/23/05
Date Completed: 3/24/05
Searcher Prep/Rev. Time: 20
Online Time: 25

Type of Search

NA#: ✓ AA#:
Interference: SPDI:
S/L: Oligomer:
Encode/Transl:
Structure#: Text:
Inventor: Litigation:

Vendors and cost where applicable

STN:
DIALOG:
QUESTEL/ORBIT:
LEXIS/NEXIS:
SEQUENCE SYSTEM: ✓
WWW/Internet:
Other(Specify):

— — — — —

Query Match	100.0%	Score 2945;	DB 6;	Length 2945;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2945;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps
1 GTCGATGCACAGCGACGCTCTCAGGCTTTGACATATTGGCGGAGCTGTGGCGACATCT 60				

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 23, 2005, 10:23:13 ; Search time 394.326 Seconds
(without alignments)
10493.597 Million cell updates/sec

Title: US-10-017-168-3

Perfect score: 639
Sequence: 1 atgttcgtcgscgacgcacac.....atcagttcggaaacctga 639

Scoring table: IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched: 4390206 seqs, 2959670667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_16Dec04.*

1: geneeqn19808.*
2: geneeqn19908.*
3: geneeqn20008.*
4: geneeqn20018.*
5: geneeqn20028.*
6: geneeqn20038.*
7: geneeqn20048.*
8: geneeqn20058.*
9: geneeqn20068.*
10: geneeqn20078.*
11: geneeqn20088.*
12: geneeqn20098.*
13: geneeqn20108.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	699	100.0	699	AAC84648
2	626.4	89.6	939	AAC84649
3	621.6	88.9	2945	AAC84647
4	537.8	76.9	9410	AAX20603
5	82.6	11.8	149158	ADP74211
6	82.6	11.8	149158	ADP74211
7	82.6	11.8	149158	ADP74211
8	82.6	11.8	149158	ADP74211
9	82.6	11.8	150071	ADP74216
10	82.6	11.8	150071	ADP74216
11	82.6	11.8	150223	ADP74201
12	82.6	11.8	150223	ADP74201
13	80.6	11.5	110000	ABX16390_2
14	78.2	11.2	628	ACH89664
15	77.6	11.1	3127	ADN12154
16	77.4	11.1	114793	ADN12154
17	77.2	11.0	3012	ADN12155
18	75.8	10.8	7720	AAAS3800
19	75.8	10.8	8298	AAK72613
20	75	10.7	16442	AAK83006

21	73.4	10.5	795	2	AAV55830	AAV55830
22	73.4	10.5	799	2	AAV55831	AAV55831
23	73.4	10.5	1925	2	AAV50924	AAV50924
24	73.4	10.5	1926	3	AAV50254	AAV50254
25	73.4	10.5	1926	4	AAV50254	AAV50254
26	73.4	10.5	1926	10	AAV50254	AAV50254
27	73.4	10.5	2580	3	AAV5454	AAV5454
28	73.4	10.5	2580	6	AAV5454	AAV5454
29	73.4	10.5	5452	2	AAV50923	AAV50923
30	73.4	10.5	8705	2	AAV50923	AAV50923
31	73.4	10.5	8705	12	AAV50923	AAV50923
32	73.4	10.5	9482	12	AAV50923	AAV50923
33	73.4	10.5	9482	12	AAV50923	AAV50923
34	73.4	10.5	10285	6	ABV51027	ABV51027
35	73.4	10.5	10285	6	ABV51027	ABV51027
36	73.4	10.5	10380	12	ADL67154	ADL67154
37	73.4	10.5	10380	12	ADL67154	ADL67154
38	73.4	10.5	10380	12	ADL67154	ADL67154
39	73.4	10.5	10516	12	ADL67150	ADL67150
40	73.4	10.5	10516	12	ADL67150	ADL67150
41	73.4	10.5	10596	2	AAQ51731	AAQ51731
42	73.4	10.5	10596	2	AAQ51731	AAQ51731
43	73.4	10.5	10596	2	AAQ51731	AAQ51731
44	73.4	10.5	10615	12	ADL67175	ADL67175
45	73.4	10.5	10774	12	ADL67153	ADL67153

ALIGNMENTS

RESULT 1	
AAC84648	AAC84648 standard; DNA; 699 BP.
ID	XX
AC	XX AAC84648;
DT	XX 11-SEP-2003 (revised)
DT	XX 20-APR-2001 (first entry)
XX	XX
DE	XX T. pallidum (CDC-2) acidic repeat protein (arp) encoding DNA.
XX	XX Treponema pallidum; acidic repeat protein; arp; immunogenic; syphilis;
KW	XX yaws; bejel; ds.
OS	XX Treponema pallidum; asp. pertenue.
XX	XX
FT	XX Key Location/Qualifiers
FT	XX CDS 1..699
XX	XX /*tag= a
PN	XX MO200077486-A2.
PD	XX 21-DEC-2000.
XX	XX
XX	XX 14-JUN-2000; 2000MO-US016425.
XX	XX
PR	XX 14-JUN-1999; 99US-0138981P.
XX	XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA	XX
XX	XX Liu H, Steiner B, Rhodes B;
PI	XX
XX	XX WPI; 2001-080711/09.
DR	XX P-PSDB; AAB8317.
XX	XX
PT	XX Detecting Treponema pallidum in blood, saliva, etc., by detecting
PT	XX formation of a complex between immunogenic peptides of acidic repeat
PT	XX protein of the bacterium and an antibody present in the biological
XX	XX sample.
PS	XX Claim 19; Fig 7; 73pp; English.
XX	XX
CC	XX The invention relates to a method of detecting presence of Treponema

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 23, 2005, 10:23:14 ; Search time 2482.36 Seconds
(without alignments)
10718.395 Million cell updates/sec

Title: US-10-017-168-3

Perfect score: 699
Sequence: 1 atcgttcgtcgcagtcagatc.....atcagttcgcgaacctga 699

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gsa1:
9: gb_gsa2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	89.6	12.8	365	9	AL160994 Leishman
C 2	85.8	12.3	293	8	AZ334000 IM0063P17
C 3	83	11.9	776	9	AG561854 Mus muscu
C 4	81.4	11.6	709	9	AG351718 Mus muscu
C 5	77	11.0	697	6	BY752869 BY752869
C 6	76.6	11.0	514	8	BZ089457 CH230-1C1
C 7	76.6	11.0	783	9	AG573654 Mus muscu
C 8	75.4	10.8	820	9	CL836578 OR CBA006
C 9	73.4	10.5	482	7	CV254879 W602411.B
C 10	73.2	10.5	752	9	AG562801 Mus muscu
C 11	72.8	10.4	650	8	AZ337339 IM0068B01
C 12	72.2	10.3	384	9	CNS01145 AL099791 Drosophila
C 13	71.8	10.3	523	8	AZ383375 IM0141E03
C 14	71.8	10.3	979	6	CD328191 AGENCOURT
C 15	71.4	10.2	419	4	BG273097 na199f06
C 16	70.8	10.1	1945	8	CC251395 CH261-157
C 17	70.6	10.1	275	7	CC641774 OGBU03TH
C 18	70.4	10.1	733	9	CC641774 OGBU03TH
C 19	70.4	10.1	852	8	BZ801844 PUGAG07TB
C 20	70.4	10.1	943	9	CG344918 OG2BH42TH
C 21	70.4	10.1	1099	9	CL509275 SA11.811
C 22	70.2	10.0	617	9	CE019059 t19T-988-
C 23	70.2	10.0	696	7	CK138139 VMJ135 V1
C 24	70	10.0	834	9	AG555306 Mus muscu

C 25	70	10.0	1197	9	AG346705 Mus muscu
C 26	69.4	9.9	427	7	CK286024 EST148746
C 27	69.4	9.9	1147	8	BZ051981 Jm16f09
C 28	69.2	9.9	699	6	CD493768 CDA07-A05
C 29	69.2	9.9	1271	9	CL492192 SAT1.564
C 30	69	9.9	430	5	BM900946 tC43C10.Y
C 31	69	9.9	986	9	AG390365 Mus muscu
C 32	68.6	9.8	494	7	CF609053 GERM001.0
C 33	68.6	9.8	516	7	CF607575 GEMMA01.0
C 34	68.6	9.8	583	5	B0847872 QOAS016.Y
C 35	68.6	9.8	1344	9	AG435591 Mus muscu
C 36	68.4	9.8	501	7	CV094178 FAMU_USDA
C 37	68.4	9.8	520	7	CV096855 FAMU_USDA
C 38	68.4	9.8	552	7	CM605345 USDA_FP.1
C 39	68.4	9.8	592	7	CV092833 FAMU_USDA
C 40	68.4	9.8	609	7	CV099895 FAMU_USDA
C 41	68.4	9.8	624	7	CV092434 FAMU_USDA
C 42	68.4	9.8	647	7	CV096641 FAMU_USDA
C 43	68.4	9.8	651	7	CV093263 FAMU_USDA
C 44	68.4	9.8	651	7	CV097413 FAMU_USDA
C 45	68.4	9.8	688	7	CV098193 FAMU_USDA

ALIGNMENTS

RESULT 1
L2644X/C
LOCUS
DEFINITION
L2644X 365 bp DNA linear GSS 16-MAR-2000
Leishmania major Friedlin cosmid L2644 t3Hyg end-sequence, similar
to U40394 U40394 Kaposi's sarcoma-associated herpes-like. N=755,
Prob=1.8e-54; TR:P93797 PHEROPHORIN-S PPECURSOR. [O] +3. ...
N=318, Prob=7.2e-37, genomic survey sequence.

ACCESSION
VERSION
AL160994.1 GI:7258621

KEYWORDS
SOURCE
ORGANISM
GSS.
Leishmania major
Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE
1 (bases 1 to 365)

AUTHORS
Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
Smith,D.F.

TITLE
JOURNAL
MEDLINE
PUBMED
A physical map of the Leishmania major Friedlin genome
Genome Res. 8 (2), 135-145 (1998)
9477341

REFERENCE
2 (bases 1 to 67)

AUTHORS
Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and
Barrell,B.G.

TITLE
JOURNAL
Submitted (14-MAR-2000) Leishmania major Friedlin genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
a1at@sanger.ac.uk

COMMENT

see http://www.ebi.ac.uk/parasites/leish.html
Details of Leishmania sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/L_major/
The c3Hyg t3Hyg primer sequence can be obtained from acc. no.
US9231.

FEATURES
source Location/Qualifiers

1..365
/organism="Leishmania major"
/mol_type="Genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="26"
/clone="cosmid L2644"

ORIGIN

Query Match 12.8% Score 89.6; DB 9; Length 365;
Best Local Similarity 58.7%; Pred. No. 4.5e-13;
Matches 155; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using BW model

Run on: March 23, 2005, 10:23:13 ; Search time 3122.69 Seconds

(without alignments)
10846.472 Million cell updates/sec

Title: US-10-017-168-3

Perfect score: 699

Sequence: 1 atgcttcgagcagcagtcacat.....atcagttcggaaacccctga 699

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenEmbl:*
2: gb_ha:*
3: gb_hcg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_ro:*
12: gb_sbs:*
13: gb_ay:*
14: gb_un:*
15: gb_vt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	699	100.0	699	6	AX068034 Sequence
2	693.4	99.2	1047	1	AF342806 Treponema
3	626.4	89.6	939	6	AX068036 Sequence
4	626.4	88.9	1287	1	AF342807 Treponema
5	621.6	88.9	1647	1	AF411124 Treponema
6	621.6	88.9	2945	6	AX068032 Sequence
7	621.6	88.9	2945	1	AF015824 Treponema
8	608.8	87.1	14268	1	AE001220 Treponema
9	601.4	86.0	1182	1	AF411126 Treponema
10	89.4	12.8	21580	2	AC111089 Mus muscu
11	88.4	12.6	165592	10	AC122020 Mus muscu
12	87.8	12.6	219771	2	BX005376 Dario rer
13	87.8	12.6	253504	10	AC102602 Mus muscu
14	87.8	12.6	266556	2	AC136868 Rattus no
15	87.6	12.5	239393	2	AC105853 Rattus no
16	87.2	12.5	269572	2	AC108552 Rattus no
17	86.8	12.4	110000	10	AE016753 Contunation (3 of
18	86.8	12.4	152435	10	AC121566 Mus muscu
19	86.2	12.3	125681	10	AC122889 Mus muscu

C 20	86.2	12.3	161905	10	AL929034	AL929034 Mouse DNA
C 21	86.2	12.3	178757	10	AC122509	AC122509 Mus muscu
C 22	86.2	12.3	196236	10	AC108780	AC108780 Mus muscu
C 23	86.2	12.3	197411	10	AC135961	AC135961 Mus muscu
C 24	86.2	12.3	203086	10	AC099948	AC099948 Mus muscu
C 25	86.2	12.3	215769	10	AC100209	AC100209 Mus muscu
C 26	85.6	12.2	192371	10	AC123857	AC123857 Mus muscu
C 27	85.6	12.2	194459	10	AC125192	AC125192 Mus muscu
C 28	85.2	12.2	143913	2	AC124827	AC124827 Mus muscu
C 29	84.6	12.1	120487	10	AC103637	AC103637 Mus muscu
C 30	84.6	12.1	163882	10	AC133904	AC133904 Mus muscu
C 31	84.6	12.1	171593	10	AC122052	AC122052 Mus muscu
C 32	84.6	12.1	179063	10	AC145590	AC145590 Mus muscu
C 33	84.6	12.1	191753	10	AC125235	AC125235 Mus muscu
C 34	84.6	12.1	199854	10	AL591208	AL591208 Mouse DNA
C 35	84.6	12.1	210595	2	AC116862	AC116862 Mus muscu
C 36	84.6	12.1	215620	10	AL928551	AL928551 Mouse DNA
C 37	84.6	12.1	217353	10	AC149611	AC149611 Mus muscu
C 38	84.6	12.1	229190	10	AC127348	AC127348 Mus muscu
C 39	84.6	12.1	234888	2	AC102097	AC102097 Mus muscu
C 40	84.6	12.1	260625	2	AC096018	AC096018 Rattus no
C 41	84.6	12.1	289751	10	AC091458	AC091458 Mus muscu
C 42	84	12.0	179886	2	AC140072	AC140072 Mus muscu
C 43	84	12.0	182272	10	AC122232	AC122232 Mus muscu
C 44	84	12.0	208880	10	AC123837	AC123837 Mus muscu
C 45	84	12.0	219568	10	AL844840	AL844840 Mouse DNA

ALIGNMENTS

RESULT 1
AX068034 699 bp DNA linear PAT 19-JAN-2001
LOCUS Sequence 3 from Patent WO0077486.
DEFINITION AX068034 GI:12329842
VERSION
KEYWORDS
SOURCE
ORGANISM
Treponema pallidum
Treponema pallidum
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.

REFERENCE
AUTHORS
TITLE
JOURNAL
Liu, H., Steiner, B. and Rhodes, B.
Compositions and methods for detecting Treponema pallidum
Patent: WO 0077486-A 3 21-DEC-2000;
THE GOVERNMENT OF THE UNITED STATES OF AMERICA (US)

FEATURES

source

CDS

1..699
/note="unnamed protein product; Subspecies: pertenue (CDC-2 strain)"
/organism="Treponema pallidum"
/mol_type="unassigned DNA"
/db_xref="taxon:160"
1..699
/codon_start=1
/transl_table=11
/protein_id="CAC24647.1"
/db_xref="GI:12329843"
/translation="MFRSDMPKNTAVRISNLERNAKQAVVIGHAGTIGLIVSLP
AAAKGIGVYQAVRVKRVKTVGKSGSOTSDGLSLSPKVPKQDVLSP
AGHVEYRDVYFDPRVLSPLSRVEVEVPKVEPASEGGEREVEVPKVEPAS
EREQGEREVEVPKVEPASEGGEREVEVPKVEPASEGGEREVEVPKVEPAS
HVSNSAPNQFRKP"

ORIGIN

Query Match 100.0%; Score 699; DB 6; Length 699;
Best Local Similarity 100.0%; Pred. No. 2.5e-145; Indels 0; Gaps 0;
Matches 699; Conservative 0; Mismatches 0;

1 ATGTTTGTCGAGTACATGTTCCCAAAACACTGCTTGAATTGAACACTAGAA 60
1 ATGTTTGTCGAGTACATGTTCCCAAAACACTGCTTGAATTGAACACTAGAA 60

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 23, 2005, 10:23:13 ; Search time 4194.86 Seconds

(without alignments)
10846.472 Million cell updates/sec

Title: US-10-017-168-5

Sequence: 1 atgcttcgagcagtcagcatt.....atcagtttcggaacctga 939

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sbs.*
12: gb_by.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	939	100.0	939	6	AX068036 Sequence
2	933.4	99.4	1287	1	AF342807 Treponema
3	858.4	91.4	1647	1	AF411124 Treponema
4	856.8	91.2	2945	6	AX068032 Sequence
5	856.8	91.2	2946	1	AF015824 Treponema
6	771.4	82.2	14268	1	AE001220 Treponema
7	721.4	76.8	1182	1	AF411126 Treponema
8	626.4	66.7	699	6	AX068034 Sequence
9	626.4	66.7	1047	1	AF342806 Treponema
10	169.4	18.0	211580	2	AC111089 Mus muscu
11	163	17.4	178757	10	AC122509 Mus muscu
12	154.4	16.4	197411	10	AC135961 Mus muscu
13	152.2	16.2	260625	2	AC096018 Rattus no
14	150.2	16.0	318930	2	AC073495 Mus muscu
15	149.6	15.9	216602	10	AC136729 Mus muscu
16	148.6	15.8	163945	2	AC116727 Mus muscu
17	148.6	15.8	217353	10	AC149611 Mus muscu
18	148.6	15.8	229190	10	AC127348 Mus muscu
19	148	15.8	165592	10	AC122020 Mus muscu

20	147.8	15.7	211161	2	AC098271 Rattus no
21	147.6	15.7	152618	10	AC127583 Mus muscu
22	145.8	15.5	197683	10	AC118476 Mus muscu
23	145.4	15.5	200574	10	BX005304 Mouse DNA
24	145.4	15.5	215745	10	AC117585 Mus muscu
25	145.4	15.5	237814	10	AC091464 Mus muscu
26	145	15.4	179206	2	AC124108 Mus muscu
27	144.8	15.4	144093	10	AC109193 Mus muscu
28	144.6	15.4	110000	2	AC120236 Rattus no
29	144.6	15.4	192371	10	AC123857 Mus muscu
30	144.2	15.4	203490	10	AC122851 Mus muscu
31	144	15.3	120487	10	AC103637 Mus muscu
32	144	15.3	171593	10	AC122052 Mus muscu
33	143.8	15.3	182272	10	AC122232 Mus muscu
34	143.8	15.3	191753	10	AC125235 Mus muscu
35	143.2	15.3	194736	10	AC126271 Mus muscu
36	143.2	15.3	205054	10	AC098712 Mus muscu
37	143	15.2	206497	10	AC084389 Mus muscu
38	142.8	15.2	152435	10	AC121566 Mus muscu
39	142.8	15.2	253504	10	AC102602 Mus muscu
40	142.6	15.2	125681	10	AC122889 Mus muscu
41	142.6	15.2	199459	10	AC125192 Mus muscu
42	142.2	15.1	178620	10	AC121865 Mus muscu
43	142.2	15.1	199854	10	AL591208 Mouse DNA
44	142.2	15.1	208880	10	AC123837 Mus muscu
45	142.2	15.1	217522	10	AC117639 Mus muscu

ALIGNMENTS

RESULT 1
AX068036 939 bp DNA 1linear PAT 19-JAN-2001
LOCUS Sequence 5 from Patent WO0077486.
DEFINITION AX068036
ACCESSION AX068036.1 GI:12329844
VERSION
KEYWORDS
SOURCE
ORGANISM
Treponema pallidum
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.

REFERENCE
AUTHORS Liu, H., Steiner, B. and Rhodes, B.
TITLES Compositions and methods for detecting Treponema pallidum
JOURNAL Patent: WO 0077486-A 5 21-DEC-2000;
THE GOVERNMENT OF THE UNITED STATES OF AMERICA (US)
FEATURES
source location/Qualifiers
1..939
/organism="Treponema pallidum"
/mol_type="unassigned DNA"
/db_xref="taxon:160"
1..939
/note="unassigned protein product; Subspecies: endemicum (Bosnia strain)"
/codon_start=1
/transl_table=1
/protein_id="CAC24648.1"
/db_xref="GI:12329845"
/translation="MTVRSDFMPKXTAVRISNLRKNAQAVVIGHAGIPGLVSLAP
AAAQGLGVYQAVRKYRRTLTSTVRKSGSTSDGLSLSLPVRVPRPQRPLSPFP
AGHTVPEYEDVIVFDPRIVSLPVSPLSRGSEVEVDVVKVEPSPASRSGSEVEVDVVKVEP
EREGREVEVDVVKVEPSPASRSGSEVEVDVVKVEPSPASRSGSEVEVDVVKVEP
ASRSGREVEVDVVKVEPSPASRSGSEVEVDVVKVEPSPASRSGSEVEVDVVKVEP
EPASRSGREVEVDVVKVEPSPASRSGSEVEVDVVKVEPSPASRSGSEVEVDVVKVEP

ORIGIN

Query Match 100.0%; Score 939; DB 6; Length 939;
Best Local Similarity 100.0%; Pred. No. 9.4e-194; Indels 0; Gaps 0;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 ATGTTTGCGCGAGTACATGTTCCCAAAACACTGCTGTGAATTGCAACTAGAA 60
db 1 ATGTTTGCGCGAGTACATGTTCCCAAAACACTGCTGTGAATTGCAACTAGAA 60

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 10:23:13 ; Search time 529.717 Seconds
(without alignments)

10493.597 Million cell updates/sec

Title: US-10-017-168-5

Perfect score: 939
Sequence: 1 atgcgttcgcgcagtcacat.....atcagtttcggaacctga 939

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: N_Geneseq_16Dec04.*
1: geneseqn19808.*
2: geneseqn19908.*
3: geneseqn20008.*
4: geneseqn20018.*
5: geneseqn20028.*
6: geneseqn20038.*
7: geneseqn20048.*
8: geneseqn20058.*
9: geneseqn20068.*
10: geneseqn20078.*
11: geneseqn20088.*
12: geneseqn20098.*
13: geneseqn20108.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	939	100.0	939	4 AAC84649	AAC84649 T. pallid
2	856.8	91.2	2945	4 AAC84647	AAC84647 T. pallid
3	637.2	67.9	9410	2 AAX20603	AAX20603 Polynucle
4	626.4	66.7	699	4 AAC84648	AAC84648 T. pallid
5	130	13.8	30191	12 ADQ97654	ADQ97654 Mouse can
6	129.4	13.8	799	2 AAV55831	AAV55831 Nucleoid
7	129.4	13.8	1926	3 AAA50254	AAA50254 Epstein B
8	129.4	13.8	1926	4 AAF82902	AAF82902 EBV tette
9	129.4	13.8	1926	10 ADK5580	ADK5580 Human her
10	129.4	13.8	2580	3 AAA75454	AAA75454 Nucleoid
11	129.4	13.8	2580	6 AAI64275	AAI64275 Epstein-B
12	129.4	13.8	5452	2 AAX90923	AAX90923 Anti-sens
13	129.4	13.8	8705	2 AAZ23778	AAZ23778 Vector ps
14	129.4	13.8	8705	12 ADM10659	ADM10659 Expressio
15	129.4	13.8	9482	12 ADP64415	ADP64415 Vector PC
16	129.4	13.8	9600	2 AAV21683	AAV21683 Vector pl
17	129.4	13.8	10285	6 ABS71027	ABS71027 pCBP-Xa-F
18	129.4	13.8	10285	6 ABS66453	ABS66453 Plasmid p
19	129.4	13.8	10330	12 ADL67154	ADL67154 Plasmid p
20	129.4	13.8	10380	2 AAZ22248	AAZ22248 Nucleoid

C 21	129.4	13.8	10477	12	ADL67152	AdL67152 Plasmid p
C 22	129.4	13.8	10516	12	ADL67150	AdL67150 Plasmid p
C 23	129.4	13.8	10561	12	ADL67148	AdL67148 Plasmid p
C 24	129.4	13.8	10596	2	AAO51731	AAO51731 Plasmid p
C 25	129.4	13.8	10596	2	AAI55650	AAI55650 Nucleoid
C 26	129.4	13.8	10596	2	AAI40348	AAI40348 Plasmid p
C 27	129.4	13.8	10615	12	ADL67175	AdL67175 Plasmid p
C 28	129.4	13.8	10774	12	ADL67153	AdL67153 Plasmid p
C 29	129.4	13.8	10921	12	ADL67151	AdL67151 Plasmid p
C 30	129.4	13.8	10961	12	ADL67149	AdL67149 Plasmid p
C 31	129.4	13.8	11006	12	ADL67147	AdL67147 Plasmid p
C 32	129.4	13.8	11059	12	ADL67176	AdL67176 Plasmid p
C 33	129.4	13.8	11693	13	ADL12379	AdL12379 Vector PC
C 34	129.4	13.8	11924	12	ADCO07395	ADCO07395 Modified
C 35	129.4	13.8	12242	12	ADCO07394	ADCO07394 Modified
C 36	129.4	13.8	16080	3	AAA59553	AAA59553 DNA clone
C 37	129.4	13.8	17753	12	ADL71910	AdL71910 Expressio
C 38	129.4	13.8	17281	12	ADN12161	ADN12161 Epstein-B
C 39	129.2	13.8	1925	2	AAV90924	AAV90924 Human gen
C 40	128.4	13.7	795	2	AAV55830	AAV55830 FlGA Inse
C 41	125.2	13.3	628	12	ACH89664	ACH89664 Human gen
C 42	120.8	12.9	26214	12	ADQ97246	ADQ97246 Mouse can
C 43	118.8	12.7	12733	6	ABK98631	ABK98631 Vector PE
C 44	118.8	12.7	12733	9	ACD13882	ACD13882 L. lactis
C 45	118.8	12.7	12739	6	ABK98592	ABK98592 Vector PE

ALIGNMENTS

RESULT 1	
AAC84649	AAC84649 standard; DNA; 939 BP.
XX	XX
AC	AAC84649;
XX	XX
XX	11-SEP-2003 (revised)
DT	20-APR-2001 (first entry)
XX	XX
DE	T. pallidum (Bosnia) acidic repeat protein (arp) encoding DNA.
XX	XX
KM	Treponema pallidum; acidic repeat protein; arp; immunogenic; syphilis;
KM	yaws; bejel; ds.
OS	Treponema pallidum; ssp. endemicum.
XX	XX
FT	Key Location/Qualifiers
FT	CDS 1..939
FT	FTag= a
XX	XX
XX	WO200077486-A2.
XX	XX
PD	21-DEC-2000.
XX	XX
PF	14-JUN-2000; 2000WO-US016425.
XX	XX
PR	14-JUN-1999; 99US-0138981P.
XX	XX
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
PI	Liu H, Steiner B, Rhodes B;
XX	XX
DR	WPI; 2001-080711/09.
XX	XX
PT	P-PSDB; AAB48318.
XX	XX
PT	Detecting Treponema pallidum in blood, saliva, etc., by detecting
PT	formation of a complex between immunogenic peptides of acidic repeat
PT	protein of the bacterium and an antibody present in the biological
XX	XX
PS	sample.
XX	XX
XX	ClaIn 19; Fig 9; 73p; English.
CC	The invention relates to a method of detecting presence of Treponema

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 23, 2005, 10:23:14 ; Search time 3334.67 Seconds
(without alignments)
10718.395 Million cell updates/sec

Title: US-10-017-168-5

Perfect score: 939
Sequence: 1 atcgttcgtcgccagtcgacat.....atcagttccggaaccctga 939

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gsa1.*
9: gb_gsa2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134.8	14.4	776	9	AG561854 Mus muscu
2	129	13.7	709	9	AG351718 Mus muscu
3	126.2	13.4	650	8	AZ337339 1M0068B03
4	124.6	13.3	1089	8	BM468944 AGENCOURT
5	122.8	13.1	623	9	CRO55317 Reverse 8
6	122.8	13.1	1110	9	CL079755 Reverse 8
7	122.8	13.1	1313	7	CF874294 trlC006xm
8	122.6	13.1	523	8	AZ383375 1M0141E03
9	122	13.0	1945	8	CC251395 CH261-157
10	121.2	12.9	1083	7	CV520573 0089P0047
11	121.2	12.9	1448	8	CC220110 CH261-92F
12	120.4	12.8	860	6	CB904993 trlC040XP
13	120.2	12.8	697	6	BY752869 Reverse 8
14	120.2	12.8	1031	9	CL492176 SAIL 564
15	120.2	12.8	1231	9	AG363910 Mus muscu
16	120	12.8	1555	9	CL073978 CH216-131
17	119.8	12.8	1560	9	CL081488 CH216-162
18	119.8	12.8	1656	9	AG430456 Mus muscu
19	119.6	12.7	602	9	CRO23579 Reverse 8
20	119.6	12.7	870	2	BE258873 60116824
21	119.6	12.7	1292	8	CC208790 CH261-26P
22	119.2	12.7	1484	8	CL079800 CH216-156
23	119.2	12.7	817	8	AZ196702 SP_1032_B
24	119.2	12.7	1546	9	AG435295 Mus muscu

25	119	12.7	935	4	B1958090 HVSMEN001
26	119	12.7	1153	8	CC294165 CH261-130
27	118.8	12.7	5107	7	TA21B11Q AL453768 T. brucei
28	118.8	12.7	1100	8	CC262597 CH261-167
29	118.8	12.7	1233	9	AG396109 Mus muscu
30	118.6	12.6	1196	8	CC298943 CH261-22C
31	118.4	12.6	1035	9	CL086207 ISB1-7A17
32	118.4	12.6	1082	8	CC300639 CH261-68U
33	118.4	12.6	1309	9	CL080508 CH216-158
34	118.4	12.6	1373	9	AG435617 Mus muscu
35	118.2	12.6	548	9	BX121682 Danto rer
36	118.2	12.6	837	9	CR074624 Forward s
37	118.2	12.6	883	9	AZ186646 SP_1007_B
38	118.2	12.6	911	8	CR116264 Forward B
39	118.2	12.6	1238	9	AG448606 Mus muscu
40	118.2	12.6	1471	9	CG748176 P042-1-A0
41	118.2	12.6	1696	9	AG382169 Mus muscu
42	118	12.6	845	9	CC578225 CH240-457
43	118	12.6	1011	6	CB905602 trlC075Xf
44	118	12.6	1011	7	CF877292 trlC075Xf
45	118	12.6	1193	9	AG349830 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS AG561854/c
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-481H05.T7, genomic survey
ACCESSION AG561854
VERSION AG561854.1 GI:48322552
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
REFERENCE 1
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC and Sequences of Library MSMg01
Unpublished
2 (bases 1 to 776)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kunita Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, Fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY : PBACE3.6
Vector : Ecoli
R.Site 1 : Ecoli
R.Site 2 : Ecoli.
FEATURES
source location/Qualifiers
1..776
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-481H05.T7"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 10:23:13 ; Search time 5280.43 Seconds
(without alignments)
10846.472 Million cell updates/sec

Title: US-10-017-168-25

Perfect score: 1182

Sequence: 1 atcgttctgcgcagtgacat.....ttctgaaaaggcgcgatga 1182

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GeneEmb1.*

1: gb_da.*

2: gb_hcg.*

3: gb_in.*

4: gb_cm.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_str.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1182	100.0	1182	1	AF411126	AF411126 Treponema
2	1085.2	91.8	14268	1	AE001220	AE001220 Treponema
3	1067	90.3	1287	1	AF342807	AF342807 Treponema
4	902	76.3	1047	1	AF342806	AF342806 Treponema
5	788	66.7	1647	1	AF411124	AF411124 Treponema
6	777	65.7	2945	6	AX068032	AX068032 Sequence
7	777	65.7	2946	1	AF015824	AF015824 Treponema
8	721.4	61.0	939	6	AX068036	AX068036 Sequence
9	601.4	50.9	699	6	AX068034	AX068034 Sequence
10	134.8	11.4	211580	2	AC111089	AC111089 Mus muscu
11	131.6	11.1	178757	10	AC122509	AC122509 Mus muscu
12	130	11.0	215769	10	AC100209	AC100209 Mus muscu
13	128.6	10.9	266556	2	AC136868	AC136868 Rattus no
14	128	10.8	199459	10	AC125192	AC125192 Mus muscu
15	126.8	10.7	125681	10	AC122889	AC122889 Mus muscu
16	124.2	10.5	165592	10	AC122020	AC122020 Mus muscu
17	123.6	10.4	199854	10	AL591208	AL591208 Mouse DNA
18	123	10.4	289751	10	AC091458	AC091458 Mus muscu
19	123	10.4	289751	10	AC091458	AC091458 Mus muscu

20	122.6	10.4	152435	10	AC121566	AC121566 Mus muscu
21	122.6	10.4	197411	10	AC135961	AC135961 Mus muscu
22	122.4	10.4	139630	10	AC133945	AC133945 Mus muscu
23	122.2	10.3	178258	2	AC109296	AC109296 Mus muscu
24	122	10.3	120487	10	AC103637	AC103637 Mus muscu
25	122	10.3	171593	10	AC122052	AC122052 Mus muscu
26	122	10.3	216602	10	AC136729	AC136729 Mus muscu
27	122	10.3	217253	10	AC149611	AC149611 Mus muscu
28	122	10.3	219771	2	BX005376	BX005376 Dantio rex
29	122	10.3	229190	10	AC127348	AC127348 Mus muscu
30	121.6	10.3	260625	2	AC096018	AC096018 Rattus no
31	121.4	10.3	182272	10	AC122232	AC122232 Mus muscu
32	121.4	10.3	208880	10	AC123837	AC123837 Mus muscu
33	121	10.2	163945	2	AC116727	AC116727 Mus muscu
34	120.6	10.2	110000	10	AB016753-2	AB016753-2 Continuation (3 of
35	120.4	10.2	318930	2	AC073495	AC073495 Mus muscu
36	119.8	10.1	154462	2	AC145149	AC145149 Homo sapi
37	119.8	10.1	191753	10	AC125235	AC125235 Mus muscu
38	119.8	10.1	196236	10	AC108780	AC108780 Mus muscu
39	119.8	10.1	197683	10	AC118476	AC118476 Mus muscu
40	119.8	10.1	203086	10	AC099948	AC099948 Mus muscu
41	119.6	10.1	215745	10	AC117585	AC117585 Mus muscu
42	119.6	10.1	211161	2	AC098271	AC098271 Rattus no
43	119.2	10.1	172853	10	AC120147	AC120147 Mus muscu
44	119.2	10.1	182777	10	AC115928	AC115928 Mus muscu
45	119.2	10.1	192371	10	AC123857	AC123857 Mus muscu

ALIGNMENTS

RESULT 1
AF411126 1182 bp DNA linear BCT 26-SEP-2001
LOCUS
DEFINITION Treponema pallidum subsp. pertenue strain CDC1 acidic repeat
protein (arp) gene, complete cds.
ACCESSION AF411126
VERSION AF411126.1 GI:15778314
KEYWORDS
SOURCE Treponema pallidum subsp. pertenue (Yaws treponeme)
ORGANISM Treponema pallidum subsp. pertenue
REFERENCE
1 (bases 1 to 1182)
Lin, H., Steiner, B.M. and Rodas, B.
Molecular characterization of the acidic repeat protein gene (arp)
of Treponema pallidum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1182)
Lin, H., Steiner, B.M. and Rodas, B.
Direct Submission
AUTHORS Submitted (17-AUG-2001) National Center for Infectious Diseases,
Centers for Disease Control and Prevention, 1600 Clifton Rd,
Atlanta, GA 30333, USA

FEATURES

source
1..1182
/organism="Treponema pallidum subsp. pertenue"
/mol_type="genomic DNA"
/strain="CDC1"
/sub_species="pertenue"
/db_xref="taxon:168"
1..1182
/gene="arp"
1..1182
/gene="arp"
1..1182
/codon_start=1
/transl_table=1
/product="acidic repeat protein"
/protein_id="AAL07373.1"
/db_xref="GI:15778315"
/translation="MPVRSDFPKRTAVESINERKKAQAQVIVGHAGIPGLVSLAP
AAAOGLGVQAVRVRRTGTGRGSGTSODGLSLASLPSRVAPAPORPDSPP
AGHTVPEYRDTVIFDDPRIVSPLSRBGGREVEDVPKVPASRBGGREVEDVPKVP
VEPASRBGGREVEDVPKVPASRBGGREVEDVPKVPASRBGGREVEDVP

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 10:23:14 ; Search time 4197.64 Seconds

(without alignments)
10718.395 Million cell updates/sec

Title: US-10-017-168-25

Perfect score: 1182

Sequence: 1 atgctctgcgcgcgcgcgcacat.....ttctgaaaaagcgcgatga 1182

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

EST: *
1: gb_esc1.*
2: gb_esc2.*
3: gb_hic.*
4: gb_esc3.*
5: gb_esc4.*
6: gb_esc5.*
7: gb_esc6.*
8: gb_gsa1.*
9: gb_gsa2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	118.2	10.0	776	9	AG561854	Mus muscu
2	116.6	9.9	709	9	AG351718	Mus muscu
3	111.6	9.4	697	6	BY752869	BY752869
4	107	9.1	650	8	AZ317339	AZ317339
5	103.4	8.7	1945	8	CC251395	CC251395
6	102.8	8.7	1560	9	CL081488	CL081488
7	101.8	8.6	1101	9	CNS0181P	CNS0181P
8	101.6	8.6	1231	9	AG363910	AG363910
9	101.4	8.6	992	9	CNS04R5Q	CNS04R5Q
10	101.2	8.6	1110	9	CL079755	CL079755
11	101.2	8.6	1448	8	CC230110	CC230110
12	101	8.5	870	2	BE258873	BE258873
13	101	8.5	1313	7	CF874294	CF874294
14	100.8	8.5	1656	9	AG430456	AG430456
15	100.6	8.5	783	9	AG573654	AG573654
16	100	8.5	548	9	BX121682	BX121682
17	100	8.5	1100	8	CC262597	CC262597
18	100	8.5	1101	8	CNS016LP	CNS016LP
19	99.8	8.4	684	8	AZ156639	AZ156639
20	99.8	8.4	974	7	CK420551	CK420551
21	99.6	8.4	1143	8	CC217764	CC217764
22	99.6	8.4	960	5	BQ429353	BQ429353
23	99.4	8.4	631	5	BU292611	BU292611
24	99.2	8.4	1107	6	CB905729	CB905729

25	99.2	8.4	1107	7	CF877385	CF877385
26	99.2	8.4	1246	9	CL058545	CL058545
27	99	8.4	523	8	AZ383375	AZ383375
28	98.8	8.4	537	9	CR080757	CR080757
29	98.8	8.4	1238	9	AG448606	AG448606
30	98.8	8.4	1484	9	CL079800	CL079800
31	98.6	8.3	432	9	BX98216	BX98216
32	98.6	8.3	872	9	CNS02GKH	CNS02GKH
33	98.6	8.3	1277	9	CG753585	CG753585
34	98.6	8.3	1471	9	CG748176	CG748176
35	98.6	8.3	1555	9	CL073978	CL073978
36	98.4	8.3	1031	9	CL492176	CL492176
37	98.4	8.3	1099	2	BF256465	BF256465
38	98.4	8.3	1196	9	AG382042	AG382042
39	98.4	8.3	1546	9	AG435295	AG435295
40	98.4	8.3	1596	9	AG382169	AG382169
41	98.2	8.3	487	9	CNS0060P	CNS0060P
42	98.2	8.3	567	8	AZ895651	AZ895651
43	98.2	8.3	883	8	AZ186646	AZ186646
44	98.2	8.3	904	9	AG424774	AG424774
45	98.2	8.3	1106	6	CB905422	CB905422

ALIGNMENTS

RESULT 1
AG561854/c
LOCUS
DEFINITION
Mus musculus molossinus DNA, clone: MSMG01-481H05.T7, genomic survey
sequence.
ACCESSION
AG561854
VERSION
AG561854.1 GI:48322552
KEYWORDS
GSS.
SOURCE
Mus musculus molossinus
ORGANISM
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE
1 Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSMG01
Unpublished
2 (bases 1 to 776)
JOURNAL
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Substitution
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22, Suenho-chou, Tsukuba, Ibaraki, Japan, 305-0856, Japan
(E-mail: hattori@gs.c.riken.jp, URL: http://hgp.gs.c.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the mouse BAC library MSMG01. For BAC library availability, please contact Kunita Abe (abe@gs.c.riken.jp).
The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyada, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@gs.c.riken.jp
PRIMERS
Sequencing: T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source
location/Qualifiers
1..776
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMG01-481H05.T7"
/sex="male"
/tissue="mixture of kidney and spleen"
/clone_lib="MSMG01 Mouse Male BAC Library"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 10:23:13 ; Search time 666.8 Seconds

(without alignments)
10493.597 Million cell updates/sec

Title: US-10-017-168-25

Perfect score: 1182

Sequence: 1 atcgttcgtgcgcagtcacat.....ttctgaaagcgcgatga 1182

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_16Dec04.*

1: Geneseq19808.*
2: Geneseq19908.*
3: Geneseq20008.*
4: Geneseq2001as.*
5: Geneseq2001as.*
6: Geneseq2002as.*
7: Geneseq2002as.*
8: Geneseq2003as.*
9: Geneseq2003as.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004as.*
13: Geneseq2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020.8	86.4	9410	2	AAX20603
2	777	65.7	2945	4	AAC84647
3	721.4	61.0	939	4	AAC84649
4	601.4	50.9	699	4	AAC84648
5	110.8	9.4	628	12	ACH89664
6	108.2	9.2	795	2	AAV55830
7	106	9.0	30191	12	ADQ97654
8	105.8	9.0	799	2	AAV55831
9	105.8	9.0	1926	3	AAV50254
10	105.8	9.0	1926	4	AAV50254
11	105.8	9.0	1926	10	AAV50254
12	105.8	9.0	2580	6	AAV50254
13	105.8	9.0	2580	6	AAV50254
14	105.8	9.0	5452	2	AAZ23778
15	105.8	9.0	8705	2	AAZ23778
16	105.8	9.0	8705	12	AAZ23778
17	105.8	9.0	9482	12	AAZ23778
18	105.8	9.0	9600	12	AAZ23778
19	105.8	9.0	10285	6	ABE66453
20	105.8	9.0	10285	6	ABE66453

C	21	105.8	9.0	10330	12	ADL67154	ADL67154 Plasmid p
C	22	105.8	9.0	10380	2	AAZ22248	AAZ22248 Nucleicid
C	23	105.8	9.0	10477	12	ADL67152	ADL67152 Plasmid p
C	24	105.8	9.0	10516	12	ADL67150	ADL67150 Plasmid p
C	25	105.8	9.0	10561	12	ADL67148	ADL67148 Plasmid p
C	26	105.8	9.0	10596	2	AAO51731	AAO51731 Plasmid p
C	27	105.8	9.0	10596	2	AAI5650	AAI5650 Nucleicid
C	28	105.8	9.0	10596	2	AAI40348	AAI40348 Plasmid p
C	29	105.8	9.0	10615	12	ADL67175	ADL67175 Plasmid p
C	30	105.8	9.0	10774	12	ADL67153	ADL67153 Plasmid p
C	31	105.8	9.0	10921	12	ADL67151	ADL67151 Plasmid p
C	32	105.8	9.0	10961	12	ADL67149	ADL67149 Plasmid p
C	33	105.8	9.0	11006	12	ADL67147	ADL67147 Plasmid p
C	34	105.8	9.0	11059	12	ADL67176	ADL67176 Plasmid p
C	35	105.8	9.0	11693	13	ADL12379	ADL12379 Vector pc
C	36	105.8	9.0	11934	12	ADQ07395	ADQ07395 Modified
C	37	105.8	9.0	12242	12	ADQ07394	ADQ07394 Modified
C	38	105.8	9.0	16080	3	AAAS9553	AAAS9553 DNA clone
C	39	105.8	9.0	17753	12	ADL71910	ADL71910 Expressio
C	40	105.8	9.0	17281	12	ADN12161	ADN12161 Eprein-B
C	41	105	8.9	8298	4	AAK72613	AAK72613 Human Imm
C	42	104.2	8.8	1925	2	AAK90924	AAK90924 Eprein B
C	43	100.8	8.5	26214	12	ADQ97246	ADQ97246 Mouse can
C	44	97.8	8.3	12733	6	ABK98631	ABK98631 Vector pc
C	45	97.8	8.3	12733	9	ACD13882	ACD13882 L. lactis

ALIGNMENTS

RESULT 1	AAX20603	standard; DNA; 9410 BP.
ID	AAX20603	
XX	AAX20603:	
AC	05-MAY-1999	(first entry)
DT	05-MAY-1999	(first entry)
XX	Polynucleotide sequence from the genome of Treponema pallidum.	
DE	Treponema pallidum infection; syphilis; Borrelia infection; animal;	
XX	enzyme production; ds.	
KW	Treponema pallidum.	
XX	OS	
XX	PN	WO9859034-A2.
XX	PD	30-DEC-1998.
XX	PF	23-JUN-1998; 98WO-US013041.
XX	PR	24-JUN-1997; 97US-0050667P.
XX	PA	(HUMAN) HUMAN GENOME SCI INC.
XX	PI	Fraser CM;
XX	WP	1999-081273/07.
XX	PT	New isolated Treponema pallidum nucleic acids - used to develop products for the detection, diagnosis, characterisation, prevention and therapy of T. pallidum infections, particularly syphilis.
XX	PS	Claim 1; Page 693-698; 1150pp; English.
XX	CC	AAZ20500-21243 represent polynucleotide sequences from the genome of Treponema pallidum. The sequences can be used for detection, diagnosis, characterisation, prevention and therapy for T. pallidum infections, particularly syphilis. They can also be used for detecting diseases related to Borrelia infections in animals, and for the production of biosynthetic products such as enzymes
XX	SO	Sequence 9410 BP; 1934 A; 2470 C; 2830 G; 2153 T; 0 U; 23 Other;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 10:23:13 / Search time 5749.51 Seconds
(without alignments)
10846.472 Million cell updates/sec

Title: US-10-017-168-23

Perfect score: 1287

Sequence: 1 atgtttgtcgcagtcgacat.....ttctgaaaaagcgcgatga 1287

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmb1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sta:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1287	100.0	1287	1	AF342807	AF342807 Treponema
2	1125	87.4	14268	1	AE001220	AE001220 Treponema
3	1067	82.9	1182	1	AF411126	AF411126 Treponema
4	933.4	72.5	1647	6	AX068036	AX068036 Sequence
5	909	70.6	1647	1	AF411124	AF411124 Treponema
6	896.4	69.7	2945	6	AX068032	AX068032 Sequence
7	896.4	69.7	2946	1	AF015824	AF015824 Treponema
8	797	61.9	1047	1	AF342806	AF342806 Treponema
9	626.4	48.7	699	6	AX068034	AX068034 Sequence
10	169.4	13.2	211580	2	AC111089	AC111089 Mus muscu
11	163	12.7	178757	10	AC122509	AC122509 Mus muscu
12	154.4	12.0	197411	10	AC135961	AC135961 Mus muscu
13	154	12.0	260625	2	AC096018	AC096018 Rattus no
14	150.2	11.7	318930	2	AC073495	AC073495 Mus muscu
15	149.6	11.6	216602	10	AC136729	AC136729 Mus muscu
16	148.6	11.5	163945	2	AC116727	AC116727 Mus muscu
17	148.6	11.5	217353	10	AC149611	AC149611 Mus muscu
18	148.6	11.5	229190	10	AC127348	AC127348 Mus muscu
19	148	11.5	165592	10	AC122020	AC122020 Mus muscu

20	147.8	11.5	211161	2	AC098271	AC098271 Rattus no
21	147.6	11.5	152618	10	AC127583	AC127583 Mus muscu
22	145.8	11.3	124244	10	AL929026	AL929026 Mouse DNA
23	145.8	11.3	197683	10	AC118476	AC118476 Mus muscu
24	145.4	11.3	200574	10	BX005304	BX005304 Mouse DNA
25	145.4	11.3	215745	10	AC117585	AC117585 Mus muscu
26	145.4	11.3	237814	10	AC091464	AC091464 Mus muscu
27	145	11.3	19206	2	AC124108	AC124108 Mus muscu
28	144.8	11.3	144093	10	AC109193	AC109193 Mus muscu
29	144.6	11.2	110000	2	AC120236	AC120236 Rattus no
30	144.6	11.2	192371	10	AC123857	AC123857 Mus muscu
31	144.2	11.2	120487	10	AC103637	AC103637 Mus muscu
32	144.2	11.2	171593	10	AC122052	AC122052 Mus muscu
33	144.2	11.2	203490	10	AC122851	AC122851 Mus muscu
34	143.8	11.2	182272	10	AC122232	AC122232 Mus muscu
35	143.8	11.2	191753	10	AC125235	AC125235 Mus muscu
36	143.2	11.1	194736	10	AC126271	AC126271 Mus muscu
37	143.2	11.1	205054	10	AC098712	AC098712 Mus muscu
38	143	11.1	206497	10	AC084389	AC084389 Mus muscu
39	142.8	11.1	152435	10	AC121566	AC121566 Mus muscu
40	142.8	11.1	253504	10	AC102602	AC102602 Mus muscu
41	142.6	11.1	125681	10	AC122889	AC122889 Mus muscu
42	142.6	11.1	199459	10	AC125192	AC125192 Mus muscu
43	142.2	11.0	178620	10	AC121865	AC121865 Mus muscu
44	142.2	11.0	199854	10	AL591208	AL591208 Mouse DNA
45	142.2	11.0	208880	10	AC123837	AC123837 Mus muscu

ALIGNMENTS

RESULT 1
AF342807 1287 bp DNA linear BCT 13-SEP-2001
LOCUS
DEFINITION Treponema pallidum subsp. endemicum strain Bosnia acidc repeat
protein (arp) gene, complete cds.
ACCESSION
AF342807 GI:15617206
VERSION
AF342807.2 GI:15617206
KEYWORDS
SOURCE
ORGANISM
Treponema pallidum subsp. endemicum
Treponema pallidum subsp. endemicum
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.

REFERENCE
1 (bases 1 to 1287)
Liu H., Steiner, B.
Acidic repeat protein (arp) gene sequence of Treponema pallidum subsp. endemicum (Bosnia strain)
Unpublished

JOURNAL
REFERENCE
2 (bases 1 to 1287)
Liu H., Steiner, B.M. and Rodas, B.
Molecular characterization of the acidic repeat protein (arp) of Treponema pallidum
Unpublished

JOURNAL
REFERENCE
3 (bases 1 to 1287)
Liu H. and Steiner, B.
Direct Submission

JOURNAL
TITLE
Submitted (25-JAN-2001) National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Rd. Mail Stop D13, Atlanta, GA 30333, USA
4 (bases 1 to 1287)
Liu H., Steiner, B. and Rodas, B.
Direct Submission

JOURNAL
TITLE
Submitted (13-SEP-2001) National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Rd. Mail Stop D13, Atlanta, GA 30333, USA
Sequence update by submitter
On Sep 13, 2001 this sequence version replaced gi:12667500.

FEATURES

REMARK
COMMENT
On Sep 13, 2001 this sequence version replaced gi:12667500.
location/Qualifiers
1..1287
/organism="Treponema pallidum subsp. endemicum"
/mol_type="genomic DNA"
/strain="Bosnia"
/sub_species="endemicum"
/db_xref="taxon:53436"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 23, 2005, 10:23:14 ; Search time 4570.53 Seconds
(without alignments)
10718.395 Million cell updates/sec

Title: US-10-017-168-23

Perfect score: 1287
Sequence: 1 atctcttcgcgcagtcgacat.....ttctgaaaaagcgcgacatga 1287

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: EST:
1: gb_est1:
2: gb_est2:
3: gb_hc:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_g881:
9: gb_g882:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	134.8	10.5	776	AG561854	Mus muscu
C 2	129	10.0	709	AG351718	Mus muscu
C 3	126.2	9.8	650	AZ337339	IM0068B03
C 4	124.6	9.7	1089	BM468944	AGENCOURT
C 5	122.8	9.5	623	CR055317	Reverse
C 6	122.8	9.5	1110	CL079755	CH216-156
C 7	122.6	9.5	1313	CF874294	tr1c006km
C 8	122.6	9.5	523	AZ383375	IM0141E03
C 9	122	9.5	1945	CC251395	CH261-157
C 10	121.2	9.4	1083	CV520573	0089P0047
C 11	121.2	9.4	1448	CC220110	CH261-92F
C 12	120.4	9.4	860	CB904933	tr1c040xp
C 13	120.2	9.3	697	BY752869	BY752869
C 14	120.2	9.3	1031	CL492176	SAIL_564
C 15	120.2	9.3	1231	AG363910	Mus muscu
C 16	120	9.3	1555	CL073978	CH216-131
C 17	119.8	9.3	1560	CL081488	CH216-162
C 18	119.8	9.3	1656	AG430456	Mus muscu
C 19	119.6	9.3	602	CR023579	Reverse
C 20	119.6	9.3	870	BE258873	601116824
C 21	119.6	9.3	1292	CC208790	CH261-26P
C 22	119.6	9.3	1484	CL079800	CH216-156
C 23	119.2	9.3	817	AZ196702	SP_1032_B
C 24	119.2	9.3	1546	AG435295	Mus muscu

25	119	9.2	935	4	BI958090	HVSMEN001
C 26	119	9.2	1133	8	CC294165	CC294165
C 27	118.8	9.2	517	7	TA21B110	AL453768 T. brucei
C 28	118.8	9.2	1100	8	CC262597	CH261-167
C 29	118.8	9.2	1233	9	AG396109	Mus muscu
C 30	118.6	9.2	1136	8	CC298943	CH261-22C
C 31	118.4	9.2	1035	9	CL086207	ISB1-7A17
C 32	118.4	9.2	1082	8	CC300639	CC300639
C 33	118.4	9.2	1309	9	CL080508	CH216-158
C 34	118.4	9.2	1373	9	AG435617	Mus muscu
C 35	118.2	9.2	548	9	BX121682	BX121682
C 36	118.2	9.2	837	9	CR074624	Forward
C 37	118.2	9.2	803	9	AZ186646	SP_1007_B
C 38	118.2	9.2	911	9	CR116264	Forward
C 39	118.2	9.2	1238	9	AG448606	Mus muscu
C 40	118.2	9.2	1471	9	CG748176	P042-1-A0
C 41	118.2	9.2	1696	9	AG382169	Mus muscu
C 42	118	9.2	845	9	CC578225	CH240_457
C 43	118	9.2	1011	6	CB905602	tr1c075xf
C 44	118	9.2	1011	7	CF877292	tr1c075xf
C 45	118	9.2	1193	9	AG349830	Mus muscu

ALIGNMENTS

RESULT 1
AG561854/c
LOCUS
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-481H05.T7, genomic survey
SEQUENCE
AG561854 776 bp DNA linear GSS 05-JUN-2004
AG561854
AG561854.1 GI:48322552
KEYWORDS
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
REFERENCE
1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC and Sequences of Library MSMg01
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 776)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC): 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.jp; URL:http://hgp.gsc.riken.go.jp/, TEL:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kunita Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center.
The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBAC3.6
R.Site 1 : ECORI
R.Site 2 : ECORI.
Location/Qualifiers
1. 776
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-481H05.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC library"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 10:23:13 ; Search time 726.034 Seconds
(without alignments)
10493.597 Million cell updates/sec

Title: US-10-017-168-23

Perfect score: 1287
Sequence: 1 atgttctgcgcagtcacat.....ttctgaaaagcgcgatga 1287

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1808:*
2: geneseqn1908:*
3: geneseqn2008:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query #	Length	DB ID	Description
1	933.4	72.5	939	4 AAC84649	AAC84649 T. pallid
2	933	72.5	9410	2 AAC20603	AAC20603 Polynucle
3	896.4	69.7	2945	4 AAC84647	AAC84647 T. pallid
4	626.4	48.7	699	4 AAC84648	AAC84648 T. pallid
5	130	10.1	30191	12 ADQ97654	ADQ97654 Mouse can
6	129.4	10.1	799	2 AAV55831	AAV55831 Nucleicid
7	129.4	10.1	1936	3 AAV50254	AAV50254 Epstein B
8	129.4	10.1	1926	4 AAF82902	AAF82902 EBV vecte
9	129.4	10.1	1926	10 AAK65580	AAK65580 Human her
10	129.4	10.1	2580	3 AAI75454	AAI75454 Nucleicid
11	129.4	10.1	5452	6 AAI64275	AAI64275 Epstein-B
12	129.4	10.1	5452	2 AAX90923	AAX90923 Anti-bens
13	129.4	10.1	8705	12 AAZ23778	AAZ23778 Vector ps
14	129.4	10.1	8705	12 ADM10659	ADM10659 Expressio
15	129.4	10.1	9482	12 AAD64415	AAZ23778 Vector ps
16	129.4	10.1	9600	2 AAV21683	AAV21683 Vector pl
17	129.4	10.1	10285	6 ABS71027	ABS71027 pCEP-Xa-F
18	129.4	10.1	10285	6 ABS64453	ABS64453 Plasmid p
19	129.4	10.1	10310	12 ADL67154	ADL67154 Plasmid p
20	129.4	10.1	10380	2 AAZ22248	AAZ22248 Nucleicid

ALIGNMENTS

C	21	129.4	10.1	10477	12	ADL67152	ADL67152 Plasmid p
C	22	129.4	10.1	10516	12	ADL67150	ADL67150 Plasmid p
C	23	129.4	10.1	10561	12	ADL67148	ADL67148 Plasmid p
C	24	129.4	10.1	10596	2	AAQ51731	AAQ51731 Plasmid p
C	25	129.4	10.1	10596	2	AAI55650	AAI55650 Nucleicid
C	26	129.4	10.1	10596	2	AAZ40348	AAZ40348 Plasmid p
C	27	129.4	10.1	10615	12	ADL67175	ADL67175 Plasmid p
C	28	129.4	10.1	10774	12	ADL67153	ADL67153 Plasmid p
C	29	129.4	10.1	10921	12	ADL67151	ADL67151 Plasmid p
C	30	129.4	10.1	10961	12	ADL67149	ADL67149 Plasmid p
C	31	129.4	10.1	11006	12	ADL67147	ADL67147 Plasmid p
C	32	129.4	10.1	11059	12	ADL67176	ADL67176 Plasmid p
C	33	129.4	10.1	11693	13	ADR23379	ADR23379 Vector pc
C	34	129.4	10.1	11924	12	ADQ07395	ADQ07395 Modified
C	35	129.4	10.1	12242	12	ADQ07394	ADQ07394 Modified
C	36	129.4	10.1	16080	3	AAV55953	AAV55953 DNA clone
C	37	129.4	10.1	17753	12	ADL71910	ADL71910 Expressio
C	38	129.4	10.1	172281	12	ADN12161	ADN12161 Epstein-B
C	39	129.2	10.0	1925	2	AAZ90924	AAZ90924 Epstein B
C	40	128.4	10.0	795	2	AAV55830	AAV55830 FLCA ins
C	41	125.2	9.7	628	12	ACH89664	ACH89664 Human gen
C	42	120.8	9.4	26214	12	ADQ97246	ADQ97246 Mouse can
C	43	118.8	9.2	12733	6	ABK98631	ABK98631 Vector pc
C	44	118.8	9.2	12733	9	ACD13882	ACD13882 L. lactis
C	45	118.8	9.2	12739	6	ABK98592	ABK98592 Vector pc

RESULT 1
AAC84649
ID AAC84649 standard; DNA, 939 BP.
XX
XX AAC84649;
AC
XX
DT 11-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX
XX T. pallidum (Bosnia) acidic repeat protein (arp) encoding DNA.
DE
XX Treponema pallidum; acidic repeat protein; arp; immunogenic; syphilis;
XX
XX Yaws; bejel; ds.
KW
XX
OS Treponema pallidum; ssp. endemicum.
XX
XX
FH Key 1. 939 Location/Qualifiers
FT CDS /*tag= a
FT
XX
XX WO200077486-A2.
XX
XX
XX 21-DEC-2000.
XX
XX
XX 14-JUN-2000; 2000MO-US016425.
XX
XX 14-JUN-1999; 99US-0138981P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Liu H, Steiner B, Rhodes B;
XX
XX WPI, 2001-080711/09.
XX
XX P-PSDB; AAB48318.
XX
XX
XX Detecting Treponema pallidum in blood, saliva, etc., by detecting
XX formation of a complex between immunogenic peptides of acidic repeat
XX protein of the bacterium and an antibody present in the biological
XX sample.
XX
XX Claim 19; Fig 9; 73p; English.
XX
XX The invention relates to a method of detecting presence of Treponema

/db_xref="taxon:168"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 10:23:14 ; Search time 3718.21 Seconds
(without alignments)
10718.395 Million cell updates/sec

Title: US-10-017-168-21
Perfect score: 1047
Sequence: 1 atgtctgtgcgcagtcgacat.....ttctgaaaagcgcgatga 1047

Scoring table: IDENTITY NUC
Gapop 10.0, Gapexc 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	89.6	8.6	365	9	L2644X	AL160994 Leishman
C 2	85.8	8.2	293	8	A2334000	A2334000 IM0063P17
C 3	83	7.9	776	9	AG561854	AG561854 Mus muscu
C 4	81.4	7.8	709	9	AG351718	AG351718 Mus muscu
C 5	77	7.4	697	6	BY752869	BY752869 BY752869
C 6	76.6	7.3	514	6	B2089457	B2089457 CH230-1C1
C 7	75.4	7.2	783	9	AG573654	AG573654 Mus muscu
C 8	75.4	7.2	820	9	CL836578	CL836578 OR CB4006
C 9	73.4	7.0	482	7	CV254879	CV254879 WS02411.B
C 10	73.2	7.0	752	9	AG562801	AG562801 Mus muscu
C 11	72.8	7.0	650	8	A2337339	A2337339 IM0068B03
C 12	72.2	6.9	384	9	CNS01145	AL099791 DIOBOPH11
C 13	71.8	6.9	523	8	A2383375	A2383375 IM0141E03
C 14	71.4	6.9	979	6	CD328191	CD328191 AGENCOURT
C 15	71.4	6.8	419	4	BG273097	BG273097 na199f06.
C 16	70.8	6.8	1945	8	CC251395	CC251395 CH261-157
C 17	70.6	6.7	275	7	COS46011	COS46011 LYEST2038
C 18	70.4	6.7	733	9	CC641774	CC641774 OGDUS35TH
C 19	70.4	6.7	852	8	B2801844	B2801844 PUGAG07TB
C 20	70.4	6.7	945	8	CG344918	CG344918 CG2BH42TH
C 21	70.4	6.7	1099	9	CL509275	CL509275 SATL 811
C 22	70.2	6.7	617	9	CE019059	CE019059 tigr-gss-
C 23	70.2	6.7	696	7	CK138139	CK138139 VRJ135 VI
C 24	70	6.7	834	9	AG555306	AG555306 Mus muscu

C 25	70	6.7	1197	9	AG346705	AG346705 Mus muscu
C 26	69.4	6.6	427	7	CK286024	CK286024 EST748746
C 27	69.4	6.6	1147	8	B2051981	B2051981 Jnt56f09
C 28	69.2	6.6	699	6	CD493768	CD493768 CD07-A05
C 29	69.2	6.6	1271	9	CL492192	CL492192 SATL 564
C 30	69	6.6	430	5	BM900946	BM900946 rc43c10.y
C 31	69	6.6	986	9	AG390365	AG390365 Mus muscu
C 32	68.6	6.6	494	7	CF609053	CF609053 GERM001.0
C 33	68.6	6.6	516	7	CF607575	CF607575 GERM001.0
C 34	68.6	6.6	583	5	BQ847872	BQ847872 OGA5016.y
C 35	68.6	6.6	1344	9	AG435591	AG435591 Mus muscu
C 36	68.4	6.5	501	7	CV094178	CV094178 FAMU_USDA
C 37	68.4	6.5	520	7	CV096855	CV096855 FAMU_USDA
C 38	68.4	6.5	552	7	CN605345	CN605345 USDA_FP.1
C 39	68.4	6.5	592	7	CV092813	CV092813 FAMU_USDA
C 40	68.4	6.5	609	7	CV098995	CV098995 FAMU_USDA
C 41	68.4	6.5	624	7	CV092434	CV092434 FAMU_USDA
C 42	68.4	6.5	647	7	CV096641	CV096641 FAMU_USDA
C 43	68.4	6.5	651	7	CV093263	CV093263 FAMU_USDA
C 44	68.4	6.5	651	7	CV097413	CV097413 FAMU_USDA
C 45	68.4	6.5	688	7	CV098193	CV098193 FAMU_USDA

ALIGNMENTS

RESULT 1
LOCUS L2644X/C
DEFINITION Leishmania major Friedlin cosmid L2644 c3hyg end-sequence, similar to U40394 U40394 Kaposi's sarcoma-associated herpes-like. . . N=755, Prob=1.8e-54; TR:P93797 P93797 PHEOPHORIN-S PRECURSOR. [0] +3. . . N=318, Prob=7.2e-37, genomic survey sequence.

ACCESSION AL160994.1 GI:7258621
VERSION 1
KEYWORDS GSS.
SOURCE Leishmania major
ORGANISM Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 365)
Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
A physical map of the Leishmania major Friedlin genome
Genome Res. 8 (2), 135-145 (1998)
98146435
9477341

REFERENCE 2 (bases 1 to 67)
Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and Barrett,B.G.
Direct Submission
Submitted (14-MAR-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrett@sanger.ac.uk and alicat@sanger.ac.uk

COMMENT see http://www.ebi.ac.uk/parasites/leish.html
Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Project/L_majior/
The c3hyg c3hyg primer sequence can be obtained from acc. no. U59231.

FEATURES
source Location/Qualifiers
1. 365

/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/Chromosome="26"
/clone="cosmid L2644"

ORIGIN

Query Match 8.6%; Score 89.6; DB 9; Length 365;
Best Local Similarity 58.7%; Pred. No. 1.5e-13;
Matches 155; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

